Shared Cancer Signal: Evidence from Cross-Training

EDCC 2022
October 18-20, 2022
Portland, OR

Earl Hubbell, Oliver Venn | GRAIL, LLC, a subsidiary of Illumina Inc.*

INTRODUCTION

- Cancer is a leading cause of death in the world, which has motivated efforts to control cancer including by early detection through screening.[1]
- Multi-cancer early detection (MCED) tests that detect cancer signals from cell free DNA (cfDNA) in the blood using sequencing assays and machine-learning classifiers are now a reality.[2,3]
  - One methylation-based MCED test covers >100,000 distinct genomic regions (17.2 Mb), covering >1,000,000 cytosine-guanine dinucleotides (CpGs).[2,3]
- Signals examined in such tests may be either:
  - Cancer-specific, limited to one particular cancer, or
  - Shared, applying to broad groups of cancers, which may allow detection of cancers that may not have been available during classifier training

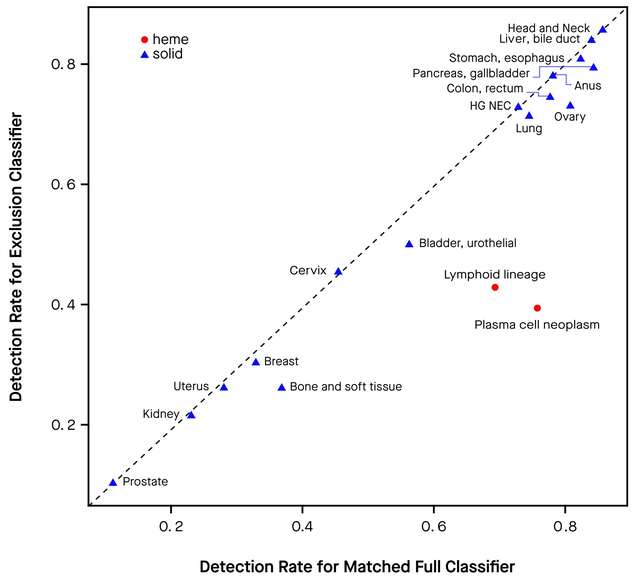
OBJECTIVE

- Examine whether a shared cancer signal across multiple cancers can be recognized in cfDNA methylation.

KEY RESULTS: SHARED CANCER SIGNAL IS FOUND ACROSS CANCERS

- Performance in solid cancers was preserved (**Figure 2**), despite exclusion from training, with a median across all solid cancers of 95% of the previous detection rate maintained.
- Performance in hematologic cancers showed a noticeable drop in detection rate of 38% and 48% for lymphoid neoplasm and plasma cell neoplasm (**Figure 2**), respectively, when excluded from training, suggesting systematic differences in signal between hematologic cancers and others.
- Performance did not decrease to that expected by chance, so some hematological cancers could share a signal with solid tumors or other hematological cancers
- Finally, training exclusively on cancers from a single anatomic region (**Figure 3**) resulted in a median of 94% of the previous detection rate in non-lung samples (lung training) or 86% of the previous detection rate in non-colorectal samples (colorectal training), indicating strong generalizability of cancer detection even when trained only on cancers from a single organ.

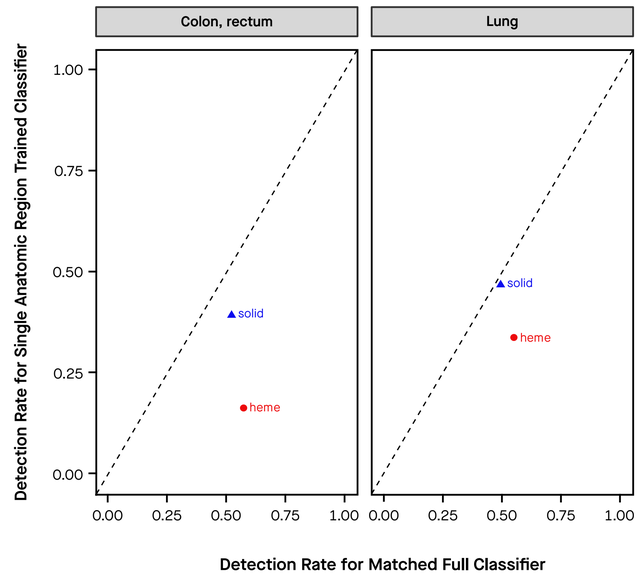
Figure 2. Single Cancer Exclusion



Detection rate for each excluded cancer using each Exclusion Classifier compared with using the Matched Full Classifier. The dashed black line represents equivalent detection rate. Hematologic cancers are indicated as red circles, solid tumors are indicated as blue triangles.

HG NEC: high grade neuroendocrine carcinomas together with all neuroendocrine tumors of the lung. Pancreas, gallbladder: includes extrahepatic bile duct cancers. Liver, bile duct: hepatocellular carcinoma and intrahepatic bile duct cancers. Bladder, urothelial: includes cancers of the renal pelvis.

Figure 3. Exclusion of All But One Cancer



Detection rate for solid and hematologic cancers using the Single Anatomic Region Classifier trained only on colon & rectum cancers (left), and only on lung cancers (right) compared with the Matched Full Classifier. In this case the Matched Classifier is trained on a reduced number of cancers to match the total number of lung or colon and rectum cancers available, respectively. The dashed black line represents equivalent detection rate.

CONCLUSIONS

- Epigenetic changes such as changes in DNA methylation have recently been noted as a "hallmark of cancer".[4]
  - Hallmarks of cancer are nearly universal properties of cancer biology
- This experiment exploring deliberate exclusion of cancers from training suggests that some epigenetic patterns found in methylation represent a shared cancer signal that allows detection of various solid cancers by MCED tests, including even those types of cancers that were not included in training.
  - Further work is ongoing to clarify a shared signal within more granular AJCC cancer types, although use of broader anatomic regions in the current analysis suggests the typical cancer type within such regions will have a shared signal.
- Consequently, MCED tests based on such a shared cancer signal have a general capability to detect multiple cancer types, including those it has not been trained to detect, when used for population-scale screening.

References

1. World Health Organization. Cancer. 2 Feb 2022. https://www.who.int/news-room/fact-sheets/detail/cancer. Accessed 19 Apr 2022.
2. Liu MC et al. *Ann Oncol.* 2020;31(6):745-759.
3. Klein EA et al. *Ann Oncol.* 2021; Sep;32(9):1167-1177.
4. Hanahan D. *Cancer Discov.* 2022;Jan;12(1):31-46.
5. Saraiya M et al. *J Natl Cancer Inst.* 2015;107(6):djv086.

Disclosures

Study funded by GRAIL, LLC, a subsidiary of Illumina Inc.* EH, OV, and AS are current or former employees of GRAIL, LLC, a subsidiary of Illumina Inc.* with equity in the company.

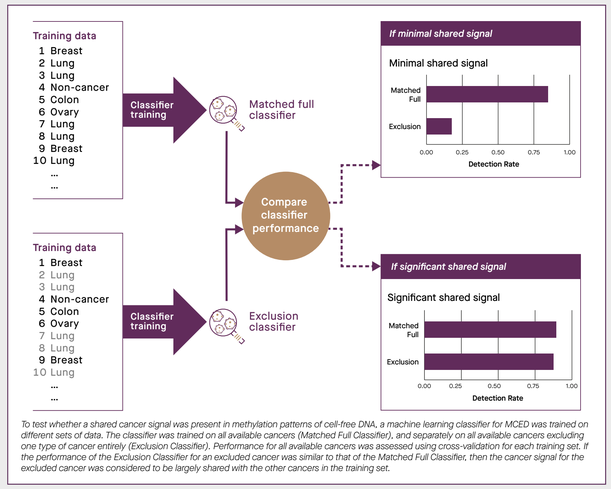
Acknowledgements

Funded by GRAIL, LLC, a subsidiary of Illumina Inc.* Graphic assistance provided by Prescott Medical Communications Group (Chicago, IL). We thank Joerg Bredno at GRAIL, LLC, a subsidiary of Illumina Inc* for helping to generate figures. *GRAIL, LLC, is currently held separate from Illumina Inc. under the terms of the Interim Measures Order of the European Commission dated 29 October 2021.

METHODS

- Using an established targeted methylation-based assay and machine-learning classification scheme, we evaluated generalizability of MCED test detection across cancers at a single high specificity.
- The MCED test detects a cancer signal by assessing alterations in cfDNA methylation patterns, and then predicts the tissue of origin of the cancer signal (cancer signal origin [CSO]).[2,3]
- We performed two types of computational experiments:
  - We excluded all cancers of a given type from the training dataset (thus, there was no trained CSO classification category for these cancers), then retrained the classifier exclusively on the other cancers
  - We trained the classifier exclusively on cancers from a single anatomic region (lung or colorectal, separately)
- Due to the reduction in the number of samples available for training in exclusion experiments, a comparable classifier using all cancers was generated using downsampling to match learning curves for training.
- Performance was evaluated on samples from cancers that were excluded from classifier training and compared to those cancers that were included in classifier training (**Figure 1**).
  - Cross-validation was used so no cancer used in training was used in evaluation

Figure 1. Cross-Training With Exclusion of Cancer Types

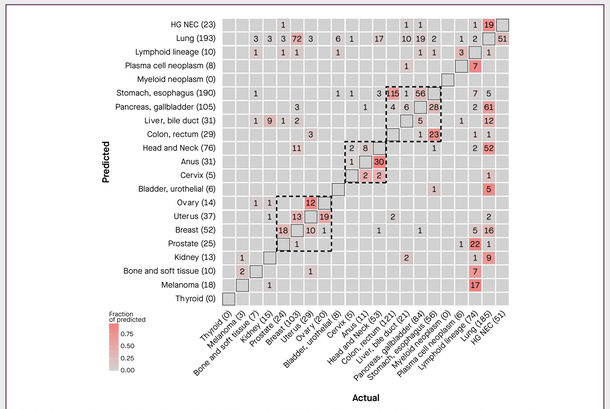


To test whether a shared cancer signal was present in methylation patterns of cell-free DNA, a machine learning classifier for MCED was trained on different sets of data. The classifier was trained on all available cancers (Matched Full Classifier), and separately on all available cancers excluding one type of cancer entirely (Exclusion Classifier). Performance for all available cancers was assessed using cross-validation for each training set. If the performance of the Exclusion Classifier for an excluded cancer was similar to that of the Matched Full Classifier, then the cancer signal for the excluded cancer was considered to be largely shared with the other cancers in the training set.

SUPPORTING DATA

- In the first experiment, the absence of a trained CSO classification category for each excluded cancer meant that no sample of the excluded cancer could have an accurate CSO prediction.
- Instead, detected cancers were observed to be classified into surrogate CSO categories with similar biological characteristics (**Figure 4**).
- The pattern of resulting mislocalizations in the confusion matrix is consistent with biological relatedness of cancers.
- Biologically related cancers formed clusters of cross-localization (**Figure 4**):
  - Cancers often driven by HPV (Head and Neck, Cervix, Anus)[5]
  - Cancers related to digestion (Colon/Rectum, Stomach/Esophagus, Liver/Bile Duct, Pancreas/Gallbladder)
  - Sex-associated cancers (Prostate, Breast, Uterus, Ovary)
- Lung and Lymphoid lineage, the most numerous solid and hematological cancers, respectively, cross localized to and were localized from multiple different cancers (**Figure 4**).
  - This may be due to the diversity of subtypes within each kind of cancer

Figure 4. Cancer Signal Origin: Predicting Localization When Cancers Were Excluded from Training



Each column contains predictions of cancer signal origin made by the Exclusion Classifier trained without that cancer. Columns and rows are ordered to reflect biological similarity so that similar cancers are together. Clusters can be observed where cancer signal origin prediction for an Exclusion Classifier returns a biologically similar cancer (outlined in dashed black boxes). The high diversity of lung cancer subtypes leads to many matches in other cancers.

HG NEC: high grade neuroendocrine carcinomas together with all neuroendocrine tumors of the lung. Pancreas, gallbladder: includes extrahepatic bile duct cancers. Liver, bile duct: hepatocellular carcinoma and intrahepatic bile duct cancers. Bladder, urothelial: includes cancers of the renal pelvis.